Using population based Kalman estimator to model COVID-19 epidemics in France: estimating the burden of SARS-CoV-2 and the effects of NPI.

Annabelle Collin[1,2,3], Boris P. Hejblum[1,4,5], Carole Vignals[1,4,5,6], Laurent Lehot[1,4,5], Rodolphe Thiébaut[1,4,5,6], Philippe Moireau[7,8], and Mélanie Prague[1,4,5,*]

[1]Inria, Inria Bordeaux - Sud-Ouest, Talence, France

[2]Bordeaux INP, IMB UMR 5251, Talence, France

[3]IMB UMR 5251, Université Bordeaux, Talence, France

[4]Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, SISTM Team, UMR 1219, F-33000 Bordeaux, France

[5]Vaccine Research Institute, F-94000 Créteil, France

[6]CHU Pellegrin, F-33000 Bordeaux, France

[7]Inria, Inria Saclay-Ile-de-France, France

[8]LMS, CNRS UMR 7707, Ecole Polytechnique, Institut Polytechnique de Paris

[*]Corresponding author

Abstract

In response to the ongoing COVID-19 pandemics caused by SARS-CoV-2, governments are taking a wide range of non-pharmaceutical interventions (NPI). These measures include interventions as stringent as strict lockdown but also school closure, bars and restaurants closure, curfews and barrier gestures. Disentangling the effectiveness of each NPI is crucial to inform response to future outbreaks. To this end, we first develop a multi-level estimation of the French COVID-19 epidemic

over a period of one year. We rely on a global extended Susceptible-Infectious-Recovered (SIR) mechanistic model of the infection including a dynamics over time for the transmission rate containing a Wiener process accounting for modeling error. Random effects are integrated by following an innovative population approach based on Kalman-type filter where the log-likelihood functional couples data across French regions. We then fit the estimated time-varying transmission rate using a regression model depending on NPI while accounting for the vaccination coverage, the apparition of variants of concern (VOC) and the seasonal weather conditions. We show that all NPI considered have a independent significant effect on the transmission rate. We additionally demonstrate a strong effect of weather condition, decreasing the transmission during the summer period, and also estimated an increased transmissibility of VOC.

1 Introduction

The World Health Organization declared the COVID-19 pandemic on 11 March 2020. This disease is caused by an infection with the SARS-CoV-2 virus. As of 30 April 2021, more than 150 million cases have been confirmed worldwide, including 3.16 million deaths. While the majority of infected cases have a mild form (upper respiratory infection symptoms) without specific needs in terms of care [1], around 3% of cases, in particular elderly, need hospitalization for treatments such as oxygenation therapy [2, 3, 4]. Among those, about 17% are severe forms (severe acute respiratory syndrome) which will need to be admitted to intensive care units (ICU) with potential need of mechanical ventilation [5]. The COVID-19 pandemic has pushed modern health care systems to a breaking point all over the world. Indeed, SARS-CoV-2 being a newly emerging pathogen, it means the entire human population is susceptible to infection. Combined with the rate of hospitalizations and ICU admissions associated with COVID-19, this leads to a surge in hospitalizations and especially ICUs needs wherever SARS-CoV-2 outbreaks arise. Due to an infectious phase starting before any symptoms are visible and a significant proportion of a- or pauci-symptomatic infections [6], the spread of SARS-Cov-2 is extremely difficult to control [7]. In response, most governments resorted to drastic public health measures, also called Non-Pharmaceutical Interventions (NPI), in order to reduce the transmission of SARS-CoV-2 among their population and consequently relax the pressure on their health-care system. Hale et al. [8] built a stringency index that helps understanding how strong the measures over time were. However, this indicator does not allow to distinguish the effectiveness of each NPI, which is crucial to inform future preparedness response plans. Because NPI all have an economic, a psychological and a social cost, it is paramount to evaluate their impact

on the transmission of SARS-CoV-2 and on the dynamic of the COVID-19 epidemic. In particular, the French government adopted the concept of a "graduated response" to the pandemic, deploying an arsenal of various NPI – some very stringent and others less so – in response to the COVID-19 national epidemic situation.

Many studies relied on epidemic models in order to either project the course of the epidemic [9], evaluate vaccine prioritization strategies [10], or retrospectively measure the NPI impact. During the beginning of the epidemics, the focus was mostly on the timing of NPI initiation [11, 12] rather than the effect. However, disentangling the effect of each NPI is a complex problem as their allocation is not randomized and depends on the epidemics state. Many approaches aggregated data from multiple countries. Some worked on regression from time series based on incidence data [13, 14, 15]. Other used semi-mechanistic models and evaluated the percentage reduction on the effective reproductive number [16, 17, 18]. We preferably work at a country level so that the effect is not confounded by the various behaviors and adherence of the population. Most of the work published at a country-level level focus on single aggregated NPI such as the Oxford COVID-19 Government Response Tracker [8], very early epidemics [19, 20], or a very limited set of interventions, see [21] for a review. Regarding France, [3, 22] quantified the effect of various NPI but based their results on the early epidemics and have a limited set of interventions. Our work rather focuses on the impact of each NPI on the transmission rate, a more valid indicator than direct epidemic curves or the reproductive number because it is independent of infected proportion of the population. Finally, existing works do not account for vaccination roll out, introduction of variants and the importance of weather while estimating the impact of NPI. The estimation of the independent effect of each of these factors leads to a challenging problem of estimation, including concerns about practical identifiability of each effect.

In this work, we propose a two-step approach. First, we estimate the transmission rate of SARS-CoV-2 and its variations in the 12 non-insular french regions over a period of more than a year - since March 2nd, 2020 until March 28, 2021. Second, we estimate using linear regression the effects of several NPI on the transmission rate while accounting for the seasonal weather conditions throughout the pandemic as well as the appearance of non-historical variants of concern (VOC) and an increasing proportion of vaccinated people. The first step consists in estimating the transmission rates in the 12 non-insular French regions. This is extremely challenging due to the sparsity and noisiness of the available data, and also because the parametric shape of the transmission rates is unknown. Using data assimilation across multiple geographical regions and coupling public data with a dynamic mechanistic model, smooth transmission rates can be estimated through a Kalman

filtering approach [23, 24] – as already used in epidemiology for COVID-19 spread [25, 26] or for other epidemics with regional variability [27]. More precisely, we develop an ingenious methodology to tackle this difficult problem, based on two important methodological innovations: (1) in the model with the introduction of a dynamics over time for the transmission rate including a Wiener process accounting for modeling error, and (2) in the way the population is integrated as we follow a new method Kalman-type filter – compatible with population approaches – presented in [28] where the log-likelihood functional – estimated using for example the Unscented Kalman Filter [29] – elegantly couples data across multiple geographical regions. These two innovations are coupled in a strategy allowing to estimate smooth transmission rates without any *a priori* on their shapes. This second step allows us to provide estimations alongside the associated uncertainties for a) the transmission rate of SARS-CoV-2 throughout the COVID-19 pandemic in France, for b) the effect of the principal NPI implemented in France on this transmission rate, and for c) the effect of the seasonal weather conditions and new VOC circulation, from the observed hospitalization data in French regions.

Section 2 presents the data and the SEIRAH model. In Sections 2.4 and 2.5, the two-step strategy to estimate the transmission rate and the regression to estimate the effects of NPI are presented. Our results are highlighted in Section 3; and their limits are discussed in Section 4.

2 Material and Methods

Open-data regarding the French COVID-19 epidemic as hospitalization data, NPI, VOC of SARS-CoV-2 and vaccination process are presented below. The evolution model of the COVID epidemic using an extended SIR type model is then presented. Finally, we describe our strategies to estimate the transmission rate using a population-based Kalman filter and to determine the impacts of NPI, seasonal weather conditions and VOC.

2.1 Available Data

2.1.1 Hospitalization Data

Hospitalization data are extracted from the SI-VIC database (Système d'Information pour le suivi des VICtimes), a governmental system created in 2016 in order to identify and follow victims in exceptional circumstances (e.g. terror attacks). Since March 18$^{\text{th}}$ 2020, the SI-VIC database provides to Santé Publique France the daily number of hospitalized patient for COVID-19, at multiple geographical scales. In this work, we focus on the 12 French non-insular regions.

An entry in the SI-VIC database corresponds to a patient to be hospitalized in connection with COVID-19. Specifically, it requires the presence of at least one of two criteria: i) a biologically confirmed diagnosis of COVID-19 (e.g. RT-PCR positive test result), or ii) a chest CT scan suggestive of COVID-19. In this analysis, we rely on the daily incident number of hospitalization ($Y^{H_{in}}$) and the total number of individual hospitalized daily (Y^{H}). The two data series are displayed in Figure 1 over a period of 391 days (from March 2$^{\text{nd}}$, 2020 to March 28$^{\text{th}}$, 2021). Of note, in order to be able to compare the magnitude of epidemics in each region, we standardized the data by the size of the population in each region: direct interpretation would be the number of incident or prevalent hospitalization for 100,000 inhabitants.

2.1.2 Non-Pharmaceutical interventions

Timing and modalities of the various NPI implemented in France over the course of the epidemics have been gathered from the french government's action summary website [30]. In France, public health intervention have been highly multi-pronged. In our analysis, we considered the following summarized NPI occurring during the first year of the epidemic in France: i) first lockdown (with two phases of relaxation/reopening as described bellow), ii) second lockdown (with one phase of relaxation/reopening as described bellow), iii) 8PM curfew, iv) 6PM curfew, v) school closure, vi) bars and restaurants closure, vii) mandatory sanitary protocols including physical distancing, hand washing, part-time remote work and mask wearing in public spaces denominated as "barrier gestures". They were considered all together as one intervention. Of note, NPI such as travel bans, enhanced testing, contact-trace-isolate, were ignored to ensure identifiability as they were either i) with a magnitude difficult to quantify, or ii) enforced in complete overlap with other NPI. In addition, partial interventions at a sub-regional level were not considered as an implementation of the measure. This resulted in a rather similar profile of interventions across regions as most of them were applied simultaneously in the 12 regions of interest. Figure 2 summarizes all 10 considered NPI over time.

The first and the second lockdowns have been disconnected due to their different modalities resulting in different behaviors, and thus possibly different impact on transmission. For example, during the first lockdown (from March 17, 2020 to May 11, 2020), all the population had to work from home – the only exceptions were for workers in vital sectors, such as medical, security, or food sectors – and outings could not exceed one hour in a one kilometre perimeter. Whereas during the second lockdown (from October 29, 2020 to December 15, 2020), working on-site was authorized when working from home was not feasible, and outings were only limited to 3 hours in

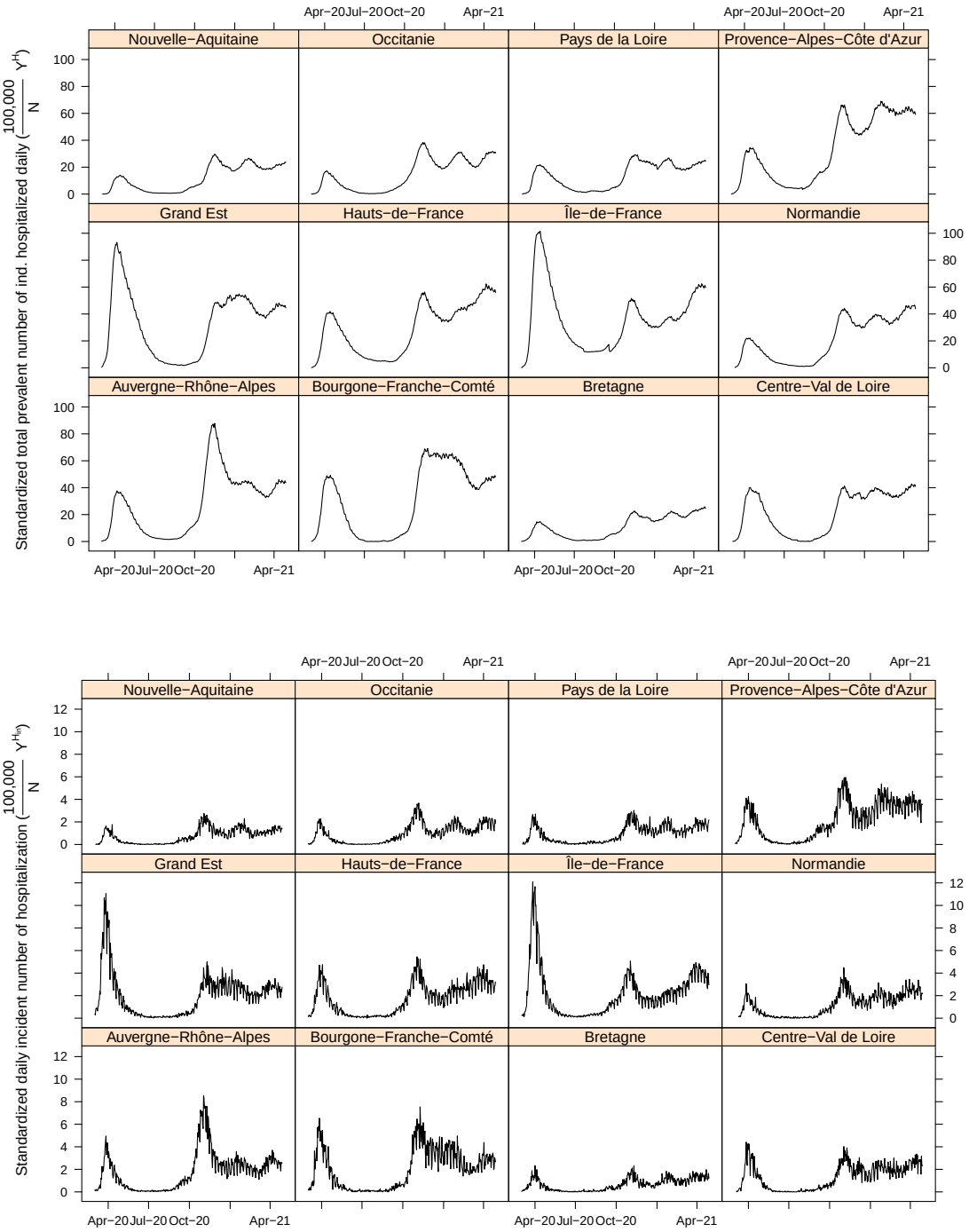

Figure 1: Top: standardized total prevalent number of individual hospitalized daily for 100,000 inhabitants ($\frac{100,000}{N} Y^H$). Bottom: standardized daily incident number of hospitalization for 100,000 inhabitants($\frac{100,000}{N} Y^{H_{in}}$).

a 20 kilometres perimeter. In addition, the end of the first lockdown was gradual and separated by the government in three official phases (phase 1: May 11, 2020 to June 2, 2020, phase 2: June 2, 2020 to June 22, 2020 and phase 3: after June 22, 2020) with many evolving measures as the authorized distance of travels, the reopening of culture places, the reopening of non-essential stores etc. A government campaign to raise awareness about barrier gestures started at the end of the first lockdown. Masks and hand washing were mandatory in many places such as public transports, schools and companies. We assume that *barrier gestures* starts on May 11$^{\text{th}}$ 2020, assuming that most of mandatory measures with a potential strong impact started to be implemented at this date. Concerning the second lockdown, a reopening of the non-essential stores happened 2 weeks prior to the end of the lockdown prior to Christmas. To account for this, we separate the second lockdown in two phases: a full lockdown until November 28$^{\text{th}}$ 2020 and a reduced lockdown afterwards.

School closure were documented as of the holiday schedule. In France, schools were also closed during the first lockdown. Furthermore, the reopening of schools from May 11, 2020 (end of first lockdown) to July 4, 2020 (end of term) was very progressive with phased schools and levels reopening and attendance rising slowly back to normal. We averaged the estimated school attendance to be 30% of school capacities during that transition. Complete closure of all of bars and restaurants happened twice over the study period: first a few days before the implementation of the first lockdown in March 2020 (as for schools closure), and along the beginning of the second lockdown in October 2020. The measure was lifted progressively in all regions after the end of the first lockdown in June 2020. At the end of the second lockdown, the measure was not lifted considering the epidemic situation was not good enough. It was still in place at the end of the study period.

Curfew measures started to be implemented October 17$^{\text{th}}$ 2020 in the 12 regions of interest in several major cities and in *Île-de-France*, with a curfew from 9PM to 6AM. It was extended on October 22$^{\text{nd}}$ to 54 departments (sub-regional administrative unit). It was suspended during the second lockdown from October 30$^{\text{th}}$ to December 15$^{\text{th}}$ 2020, when a new one was implemented at 8PM at a national level. From January 2$^{\text{nd}}$ to January 12$^{\text{th}}$ 2021 starting hour was changed to 6PM progressively at departmental level until January 16$^{\text{th}}$ when it switched to 6PM for 12 regions of interest. Finally on March 20$^{\text{th}}$ 2021, starting hour was changed again nationally to 7PM. To model the measures we considered only two variables regrouping curfews starting at 9PM or 8PM and those starting at 6PM or 7PM, and we set the value to 1 only when a whole region was under the curfew. Assuming that curfews and lockdowns induce different behaviors, curfews are not considered included in lockdowns, even if it is forbidden to go out in the evening during lockdowns. Similarly, 6PM curfew and 8PM curfew were considered as different intervention instead of nested

ones as they are likely to induce different behaviors.

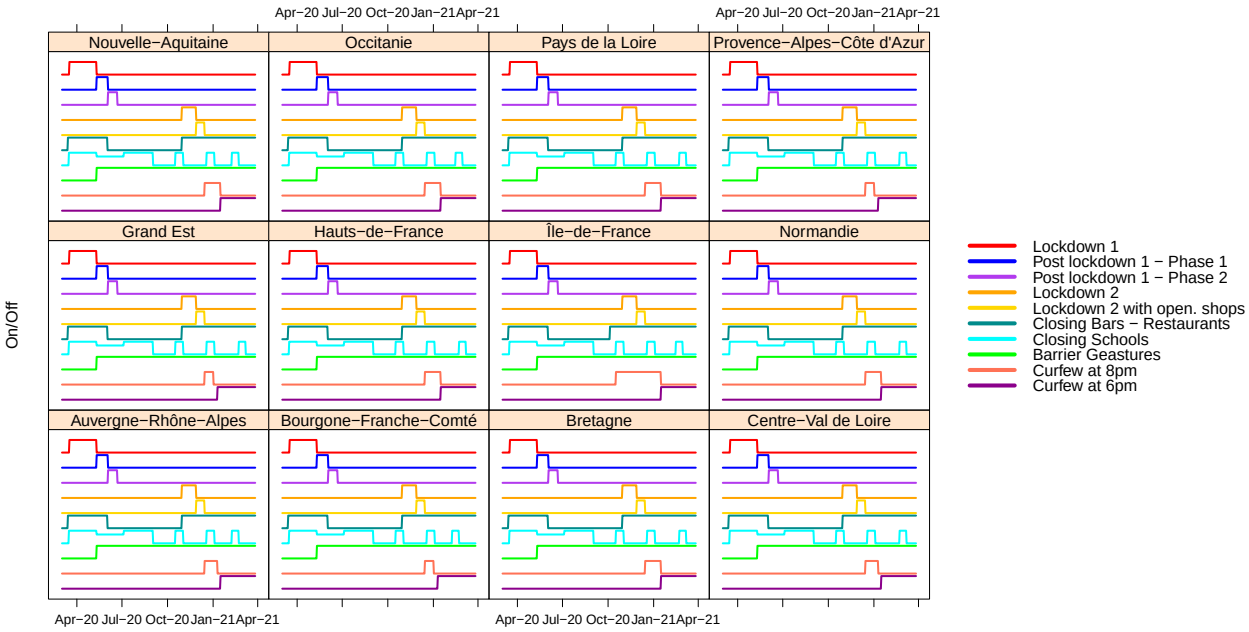

Figure 2: Major NPI at the regional level

2.2 Other exogeneous variables: weather, VOC, vaccination coverage

2.2.1 Weather conditions

The role of weather conditions on the transmission of the SARS-CoV-2 remains disputed, and early publications were criticized for their inconsistency in results [31]. Nonetheless, the potential impact of both temperature and humidity on aerosolized and fomite transmission routes based on comparisons with other respiratory infections is based on sound mechanistic arguments [32]. Additionally, with the Northern hemisphere going through a second winter season during the pandemic, evidence of the association of transmission with seasonal trends of temperature and humidity appears more robust in recent publications [32, 33]. Daily weather data – namely temperature in Celsius degrees (T), relative humidity in percentage (RH) and absolute humidity in g.m^{-3} (AH) - measured by meteorological stations were extracted from the National Oceanic and Atmospheric Administration database using `R` package *worldmet*. Regional daily means were estimated with all stations located in the region or at less than 10 km. To account for variations of population density across a region, a weighting was used based on population living in a 10km buffer around each station, giving more

importance to weather conditions around densely populated areas. We use the PREDICT index of weather transmissibility of COVID-19 (IPTCC), as defined by Bukhari and Jameel [34] and Roumagnac et al. [35]:

$$IPTCC = 100\, e^{-\frac{1}{2}\left[\frac{(T-7.5)^2}{196}+\frac{(RH-75)^2}{625}+\frac{(AH-6)^2}{2.89}\right]}.$$

As for interpretation, this rough index ranges from 0% to 100%, the smaller the less condition are favourable for COVID-19 transmissions. For France, we observe a seasonality of IPTCC being small during summer months and higher otherwise, with a north-east/south-west gradient.
From this index, we created a weather variable by normalizing (min-max range equal to 1), subtracting the global average value and by reversing it, see black curves in Figure 3. Finally, to focus on the seasonal variation, a *loess* smoothing with a span of 0.2 was applied before use leading to a smooth weather variable denoted by W in that follows. Resulting seasonal variations of this variable for each of the 12 regions of interest over the study period are shown on Figure 3 (red curves). The lower the value, the closer temperature and humidity conditions were close to optimal transmission conditions defined by [34]. Again two period appear clearly, summer with higher values of this weather variable and winter with lower values. This variable is denoted $W_i(t)$ in the following. When taking W=0, one consider the global average value over all French regions over the study period.

2.2.2 VOC of SARS-CoV-2

Among all the variants of the historical SARS-CoV-2, some are classified by national and international health authorities as VOC as they impact transmissibility or virulence, or decrease the effectiveness of measures [36]. Since January 2021, French health authorities conduct surveys to estimate the prevalence of three VOC: 20I/501Y.V1 (Alpha), 20H/501Y.V2 (Beta) and 20J/501Y.V3 (Gamma). The survey of VOC Delta started after this study. We therefore included the cumulative proportion of cases infected by one of these VOC as a potential covariate explaining the transmission. We used data from two cross-sectionnal "flash" surveys performed on January 7-8, 2021 and January 27, 2021, see [37, 38] and the weekly estimation of the VOC circulation provided by the SI-DEP database at the regional level from February 12, 2021 to March 28,2021 [39].

Between January 8, 2021 and February 12, 2021, the estimated proportion of the sum of the three VOC have increased from a national mean of 3.3% to a national mean of 46%. To complete the missing data, we assume that the proportion before January 8, 2021 equals to 0% and that the evolution is linear between January 8 and January 27, 2021, and then between January 27 and

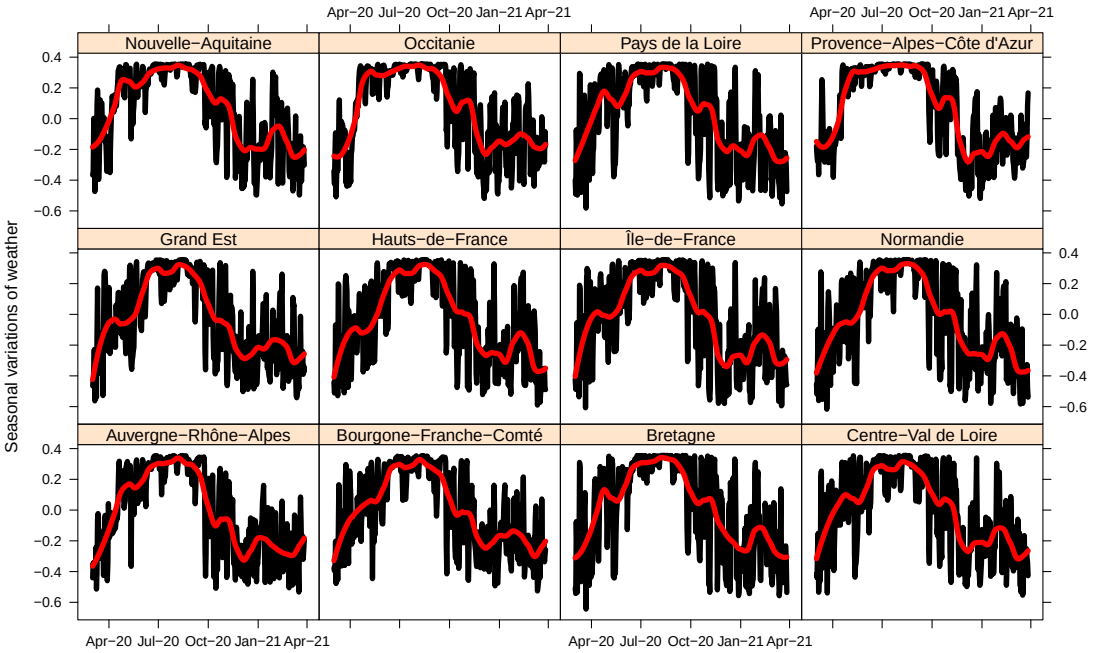

Figure 3: Weather variable modeling the seasonal weather conditions of the 12 regions of interest in black and after smoothing in red (denoted by W in this paper). The highest, the less condition of temperature and humidity are favourable to COVID transmission.

February 12, 2021. Of note, a logistic and an exponential growth were also tested with no significant changes in conclusions (results not shown). With no data reported for *Bourgogne-Franche-Comté* region on January 27, 2021, only one slope was estimated on the time window. This variable is denoted $VOC_i(t)$ in the following. Representations of the proportion of VOC in each region over time are given in Section 1 of Web Supplementary Material.

2.2.3 Vaccination process

Vaccination started in France on the December 27, 2020. Three COVID vaccines were at that time authorized: BNT162b2 mRNA (Pfizer), ChAdOx1 nCoV-19 (AztraZeneca) and mRNA-1273 (Moderna). To take into account the vaccination process, we used the VAC-SI database [40] which provides the cumulative percentage of the population who have been vaccinated with at least one dose of vaccine over time. Vaccination started on January 5, 2021 and was first targeted to the elderly 75+. The proportion of vaccinated individuals, denoted V in the following, ramped up to about 12% by the end of study period. Representations of the vaccination coverage in the population over time are given in Section 1 of Web Supplementary Material.

2.3 Model of the epidemics

The mechanistic model: We model the evolution of the COVID epidemic using an extended SEIR type model [41]. Indeed, we use a SEIRAH model adapted from [41, 42] where the population of size N is divided into 5 compartments: susceptible S, latent exposed E, symptomatic infectious I, asymptomatic/pauci-symptomatic infectious A, hospitalized H, removed R (i.e. both recovered and deceased), see Figure 4. The number of vaccinated people denoted by V is assumed to be known, see Section 2.2.3. The dynamics of such model is given by

$$\begin{cases} \dot{S} = -\underline{b}\left(1-\dfrac{V}{N}\right)\dfrac{S(I+\alpha A)}{N} \\ \dot{E} = \underline{b}\left(1-\dfrac{V}{N}\right)\dfrac{S(I+\alpha A)}{N} - \dfrac{E}{D_E} \\ \dot{I} = \dfrac{r_E}{D_E}E - \dfrac{1-r_I}{D_Q}I - \dfrac{r_I}{D_I}I, \\ \dot{R} = \dfrac{r_I I + A}{D_I} + \dfrac{H}{D_H} \\ \dot{A} = \dfrac{1-r_E}{D_E}E - \dfrac{A}{D_I} \\ \dot{H} = \dfrac{1-r_I}{D_Q}I - \dfrac{H}{D_H} \end{cases} \tag{1}$$

where $\alpha, r_E, D_E, r_I, D_I, D_Q, D_H$ are time-independent parameters described in Table 1 while b is a function of time modeling the disease transmission rate over time.

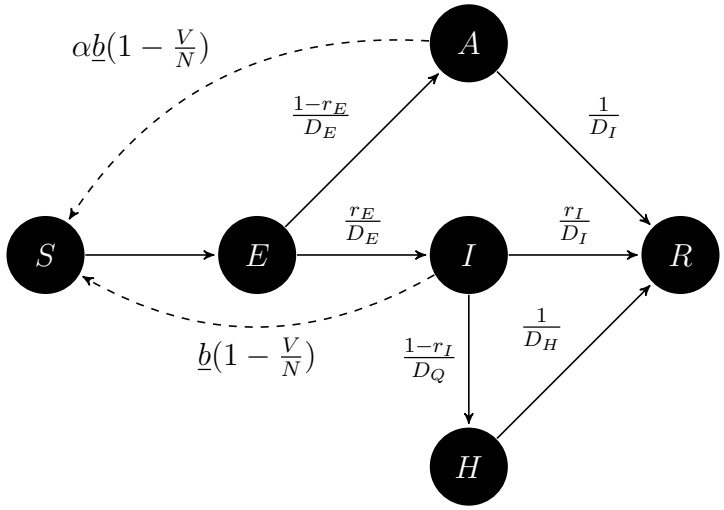

Figure 4: SEIRAH model representation – adapted from [41, 42]

The observation model: The two quantities Y^H and $Y^{H_{in}}$ relate to the solutions of System (1) respectively as for all region $i = 1, \ldots, 12$, for all observation time in days $j = 1, \ldots, 391$: $Y_{ij}^H = H_i(j) + \epsilon_{ij}^H$ and $Y_{ij}^{H_{in}} = \frac{(1-r_I)}{D_Q} I_i(j) + \epsilon_{ij}^{H_{in}}$, in which ϵ_{ij}^H and $\epsilon_{ij}^{H_{in}}$ represent normally distributed constant measurement errors.

Effective reproductive number and attack rates: When individuals are homogeneous and mix uniformly, the effective reproductive ratio $R_{\text{eff}}(t)$ is defined as the mean number of infections generated during the infectious period of a single infectious case at time t. In this model, the effective reproductive ratio can be written as a function of model parameters (see Section 3 of Web Supplementary Material for details), namely

$$R_{\text{eff}}(t) = \underline{b}(t)\Big(1 - \frac{V}{N}\Big)\frac{S(t)}{N}\left(D_I\alpha(1-r_E) + \frac{D_I D_Q r_E}{(1-r_I)D_I + r_I D_Q}\right). \tag{2}$$

When neglecting the deaths, the proportion of infected individuals assuming no waning immunity – also called attack rates – among the population in each region at a given date is given by:

$$\frac{E+I+R+A+H}{N} = 1 - \frac{S}{N}.$$

Parameter	Interpretation	Value
$\underline{b}$	Transmission rate of infectious cases	Region Specific - **Estimated**
r_E	Ascertainment rate	0.844 [43]
r_I	Non hospitalized rate	0.966 [4]
α	Ratio of transmission between A and I	0.55 [44]
D_E	Latent (incubation) period (days)	5.1 [45]
D_I	Infectious period (days)	5 [46]
D_Q	Duration from I onset to H (days)	11-D_E = 5.9 [47]
D_H	Hospitalization period (days)	18.3*
N	Population size	Region Specific

Table 1: Model parameters for the SEIRAH model, and associated values. *Computed using the correlation between the data $Y^{H_{in}}$ and Y^H when considering region data, see Section 2 of Web Supplementary Material.

2.4 A population-based Kalman filter to estimate the transmission rate

Framing the problem. In order to sustain the modeling choices behind the disease transmission $\underline{b}$ and more generally the selected inference procedure, we used Kalman-based estimation strategies [23, 24] applied to the dynamics (1). We define the global transmission rate $t \mapsto b(t)$, which accounts for the proportion of susceptible removed from the system thanks to vaccination:

$$b(t) \overset{\text{def}}{=} \underline{b}(t)\left(1 - \frac{V(t)}{N}\right).$$

We propose to estimate b and ultimately to retrieve $\underline{b}$ using an estimation of vaccinated people V, see Section 2.2.3. In each region, we then introduce a dynamics for b of the form

$$\mathrm{d}b^i(t) = g^i(t)\mathrm{d}t + \mathrm{d}\nu^i(t), \tag{3}$$

where ν^i consists in a Wiener process such as for all $t, s \geq 0$, $\nu^i(t) - \nu^i(s) \sim \mathcal{N}(0, (t-s)\sigma_\nu)$ with σ_ν^2 is known and constant in all regions and g^i is a function describing the evolution of the global transmission rates b^i in each region.

After discretization using forward Euler time-scheme with small-enough time-step δt, we end up with a discrete-time dynamical system applied to the variable $x = (E, I, R, A, H)^\intercal \in \mathbb{R}^5$, for each

region $i = 1, \ldots, 12$:

$$z^i_{n+1} \overset{\text{def}}{=} \begin{pmatrix} x^i_{n+1} \\ b^i_{n+1} \\ \theta^i_{n+1} \end{pmatrix} = \begin{pmatrix} x^i_n + \delta t\, f(x^i_n, b^i_n, \theta^i_n) \\ b^i_n + \delta t\, g^i(t_n, \theta^i_n) \\ \theta^i_n \end{pmatrix} + \begin{pmatrix} 0_5 \\ 1 \\ 0_{N_p} \end{pmatrix} \nu^i_n, \tag{4}$$

where f accounts for the dynamics of (E, I, R, A, H) in (1) while S is reconstructed afterwards by using $S = N - (E + I + R + A + H)$ in each region. In the time discrete system, $(\nu^i_n)_{n\geq 0}$ now represent independent random variables, normal with 0 mean and a variance equal to $\sigma^2_\nu \delta t$. Moreover, if constant parameters have to be estimated, the vector $\theta \in \mathbb{R}^{N_p}$ gathers all of them. For the estimation in the following, we transform the variable z to account for biological constraints. Remarking that all state variables are positive and bounded by N the total population size of the region, we transform the state variable using $x \mapsto \text{logit}(x/N)$. We also apply similar transformation to $b \mapsto \text{logit}(b/max_b)$. Of note, state and transmission rate variables are more likely to have Gaussian distribution in the transform space. After calibration, we fixed $max_b = 1.5$ and checked that other values did not substantially changed the results (result not shown).

Population approach. To perform the estimation, we decide to rely on an extension of the classical Unscented Kalman filter (UKF) [48, 29, 23]. The particularity in this application, as described in 2.4, is that multiple series of data are observed jointly in multiple regions as we observe multiple realizations of the same epidemics. In order to take into account in our Kalman estimation that parameters in different regions are correlated as in a population approach, we follow a recently proposed population-based Kalman formulation [28]. As in mixed-effect models [49], each initial uncertainty variable z^i_0 is assumed to be randomly distributed around a common population intercept z^{pop}_0 with a Gaussian distribution of unknown covariance $\boldsymbol{Q}_0$, namely:

$$z^i_0 \sim_{\text{i.i.d.}} \mathcal{N}(z^{\text{pop}}_0, \boldsymbol{Q}_0).$$

When we treat the population intercept as the empirical mean over the population members in the construction of the objective function, we recover a classical filtering problem [23] on the aggregated variable $\mathbf{z} = (z^1 \cdots z^{N_r})^\intercal$. The only difference is the formulation of the initial covariance prior $\hat{\boldsymbol{P}}_0$ which couples the observations across regions and can be written as:

$$\hat{\boldsymbol{P}}_0^{-1} = \frac{1}{N_r^2} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \otimes \boldsymbol{M} + \left[\mathbb{1}_{N_r} - \frac{1}{N_r} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \right] \otimes \hat{\boldsymbol{Q}}_0^{-1},$$

where $\otimes$ indicates a Kronecker product, $\hat{\boldsymbol{Q}}_0$ is a prior of $\boldsymbol{Q}_0$ and $\boldsymbol{M}$ a small penalization matrix that guarantees the overall invertibility of $\hat{\boldsymbol{P}}_0$. As a consequence, the matrix $\hat{\boldsymbol{P}}_0$ is not block diagonal with respect to the region i and thus all the region dynamics are coupled. The resulting time-discrete Kalman estimator couples all the regions together to produce a population-based estimation. Note that in such strategy, forcing a variable to be constant in the population is possible by simply choosing $\hat{\boldsymbol{Q}}_0$ such that $\mathrm{Tr}(\hat{\boldsymbol{Q}}_0^{-1})$ is very small with respect to $\mathrm{Tr}(\boldsymbol{M})$. Conversely, $\mathrm{Tr}(\boldsymbol{M})$ small with respect to $\mathrm{Tr}(\hat{\boldsymbol{Q}}_0^{-1})$ in a large population $N_r \gg 1$ will encourage each region to remain decoupled from each others. Starting from a given prior knowledge, our Kalman implementation uses the available measurements $(Y_{ij}^H, Y_{ij}^{H_{in}})_{1\leq i\leq 12, 0\leq j\leq 391}$, to compute recursively in time the following estimates:

$$\hat{\boldsymbol{z}}_n \simeq \mathbb{E}(\boldsymbol{z}_n | (Y_{ij}^H, Y_{ij}^{H_{in}}), 1 \leq i \leq 12, 1 \leq j \leq n), \quad 0 \leq n \leq 391,$$

and

$$\hat{\boldsymbol{P}}_n \simeq \mathbb{C}\mathrm{ov}(\boldsymbol{z}_n - \hat{\boldsymbol{z}}_n | (Y_{ij}^H, Y_{ij}^{H_{in}}), 1 \leq i \leq 12, 1 \leq j \leq n), \quad 0 \leq n \leq 391.$$

Then exploiting that $\hat{\boldsymbol{z}}_n$ gather the augmented state $(x_{n+1}^i, b_{n+1}^i, \theta_{n+1}^i)$, $1 \leq i \leq 12$ of the 12 regions, a simple post-processing over the regions provides estimations of b and the state variable. We refer to [28] for further details about this Kalman-based population approach.

Estimation strategy. Although, we want to inject as few information as possible on the parametric shape of b, we stress out that ν defined in (3) is a time-dependent function which hence can encompass the complete dynamics of b while choosing $g \equiv 0$ to avoid any *a priori* on the NPI effects. However, to avoid overfitting, our objective is to disentangle the latent trajectory of b with possible noise from data, which can be clearly visualized in Figure 1. We adopt a 3-step approach described bellow consisting in smoothing the trajectory of b:

1. Estimate a reasonable prior for the initial transmission rate b before the start of any NPI. We use data before the first lockdown (10 days available) and assume the transmission rate $b^i(t) = b_{init}^i$ for $t = 1 \cdots 10$ days to be a constant. In other words, we apply the population Kalman filter estimation described above with θ from Equation (4) being reduced to b_{init}.

2. Estimate the shape of b with prior on initial value but without any prior on the dynamics. We set the initial value of b^i for times $t = 1 \cdots 10$ days to b_{init}^i and take $g(t) \equiv 0$ such that Equation (3) rewrites $\mathrm{d}b(t) = \mathrm{d}\nu(t)$. We apply the population Kalman filter estimation described above. Of note, the parameters vector θ from Equation (4) is now empty. As there is no information on b, the model error is very important and could lead to an over-fitting of the

data. Then we build a prior for the dynamics of b by fitting a parametric shape based on sum of logistic functions on the weighted average trajectories of b over all regions with least-square method. Then, we build a prior for the dynamics of b by fitting a parametric shape based on sum of logistic functions on the weighted average trajectories of b over all regions with least-square method. We choose logistic functions as they seem well adapted to model variations due to for example stiff lockdowns or smooth unlocks. We set the number of logistic functions from the number of main changes of variations of the weighted average trajectories of b. This function will represent a prior g on the dynamics.

3. Estimate the shape of b with prior on initial value and informative prior on the dynamics. We set the initial value of b^i for times $t = 1 \cdots 10$ days to b^i_{init} and take $\mathrm{d}b(t) = g(t)\mathrm{d}t + \mathrm{d}\nu(t)$ such as in Equation (3). We apply the population Kalman filter estimation described above. As the dynamics of b still includes a modeling noise, the shape of b could be different from the prior transmission rate defined from Step 2. This is the results of this final analysis which are used for further description of the effect of NPI.

2.5 Explanatory model for the transmission rate

Mixed effects model Using the function b obtained with population Kalman filter as described in Section 2.4, one can determine the impacts of NPI, seasonal weather conditions and VOC on the transmission rate given by $\underline{b} = \frac{b}{(1-V/N)}$. We followed the example of Flaxman et al. [18] - in considering interventions effects to be multiplicative. Therefore, we fitted a linear mixed effect model on the logarithm transformation of the transmission. The model equations are given in Section 4.2 of Web Supplementary material. It consists in an intercept, which represents the average transmission of COVID-19 across all region without NPI without VOC circulation and for an average French weather condition (W=0), the sum of effect of the 10 NPI described in 2.1.2, the effect of weather and the effect of percentage of VOC. It is known that the transmission of SARS-CoV-2 is very different indoors and outdoors. Thus, we also added an interaction effect between bars and restaurants closure and the weather, accounting for the opening of outdoor sitting areas. Of note, this is the only interaction tested to avoid overfitting. We finally added random effects to account for heterogeneity between regions. We added a random intercept and random slopes for the effect of first, second lockdown and 6PM curfew as they may be highly variable between regions. We assume a full covariance matrix for the random effects, allowing effects to be correlated together in each region. In particular, we believe that they may be impacted by various factors not accounted for in the epidemic model, such as the density, the age pile or the urbanization of

the region. Note that we added a 7 days delay to the lockdowns, which can be interpreted as a necessary time to allow people to organize (implementation of working from home, childcare etc ...). This choice has been motivated by the observed 7 days delays in the transmission rates obtained by Kalman filters and will be discussed. To ease interpretation of the estimated effects, parameters were transformed back and expressed as a decrease or an increase in transmission in percentage by applying the function $x \mapsto 100(e^x - 1)$. Classical 95% confidence intervals were obtained using $100(e^{x+/-1.96SE(x)} - 1)$, were SE is the standard error obtained from the regression.

Interpretation of the impact of seasonal weather conditions To facilitate the understanding, we calculated some NPI effects during the summer and winter periods. To do so, we replace in the results W by its summer (resp. winter) average from June 21st, 2020 to October 21st, 2020 (resp. before June 21st, 2020 and after October 21st, 2020) over all regions.

Basic reproductive number The intercept of the regression presented above represents the mean transmission rate over all region when there is no NPI in place, no VOC and the weather condition is taken to be the average weather condition over a year in France. Thus, injecting it in (2) directly provides the basic reproductive number and 95% confidence intervals can be computed using the standard error of this parameter.

3 Results

Estimation of the transmission rate using a population-based Kalman filter Step 1 of estimation provided the initial values for the transmission rate and exhibited quite similar values between regions (average 0.78 sd 0.012, see Section 4.1 of Web Supplementary Material for the values at the regional level). Although higher values are found for the regions with the higher first wave, the small variability between regions may indicate that the magnitude of the first wave was not fully driven by a higher transmission rate, but also by different initial epidemic states (i.e. number of exposed and infectious cases resulting in differences in virus introduction timelines or the occurrence of super-spreading events [50]). In Step 2, The transmission rates obtained without any apriori are given in Figure 5-Top-Left for all the regions. As there is no information on b, the model error is very important and lead to an over-fitting of the data. In particular, oscillations of 0 mean and of period close to 7 days are visible for many regions. This is related to a known under-reporting during week-ends. The weighted average trajectory of b is displayed in Figure 5-Bottom-Left. We approximated the function g by a sum of 7 logistic functions. Result of the least-square fitting is

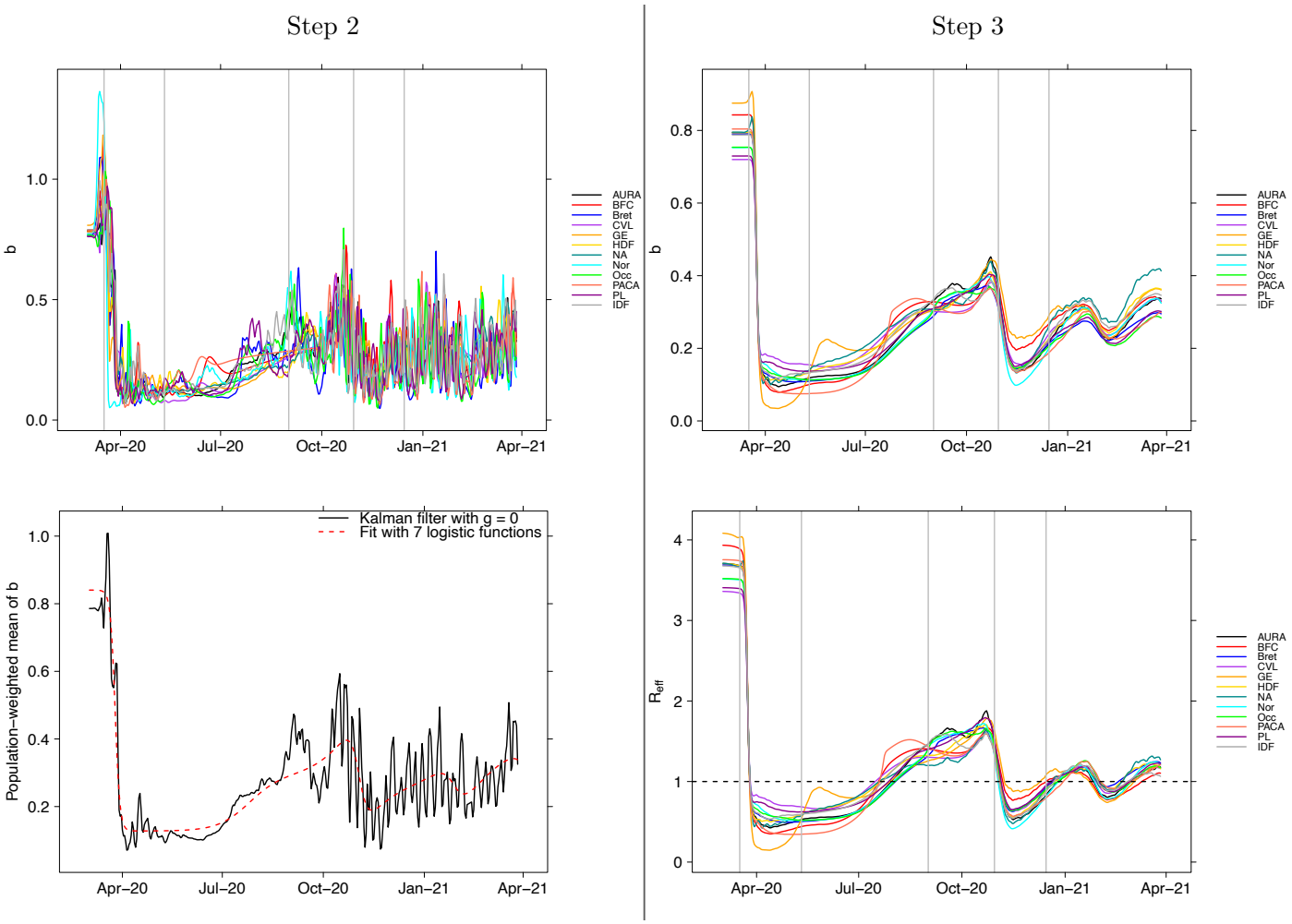

Figure 5: First column (Step 2): Top - Time evolution of b in the 12 french regions with $g = 0$ (no *apriori*). Bottom - Mean value of b over time (black line) obtained at Step 2 fitted using 7 logistic functions (dashed red line). Second column (Step 3): Top - Time evolution of b in the 12 french regions when g corresponds to 7 logistic functions obtained at Step 2 (with $\nu \neq 0$). Bottom - Time evolution of R_{eff} in the 12 french regions when g corresponds to 7 logistic functions (with $\nu \neq 0$). If R_{eff} is superior to 1 (see the horizontal dashed black line), the infection will spread. The first (resp. second, third, fourth, fifth) vertical gray line corresponds to the first day of the first lockdown (resp. the last day of the first lockdown, the start of the academic year, the first day of the second lockdown, the last day of the second lockdown).

displayed in dashed line and constitute the prior for the following. Finally, in Step 3, the transmission rates b are obtained and displayed in Figure 5-Top-Right. As the dynamics of b still includes a modeling noise, the shape of b is different from the prior transmission rate defined from Step 2 but still smoother.

Effective reproductive number The resulting effective reproductive ratio R_{eff} is given at the regional level in Figure 5-Bottom-Right. It starts from a value ranging between 3.5 and 4 in all regions. This is partially driven by a winter-like condition during the first days of March 2020. The basic reproductive number will be later computed adjusting for weather. The time variation of R_{eff} shows that it rapidly gets under the critical value of one after the first lockdown initiation and then oscillate around this value.

Attack rates The attack rates represents one intrinsic feature of interest in any epidemic model (as described in Section 2.3). On top of adding an insight on the epidemics dynamics by adding extra knowledge on the number of possible hidden / unmeasured cases, attack rates are good way to validate how realistic the model fitting is. Figure 6 represents the attack rate at various important times: at the end of first lockdown (May 11, 2020), October 5, 2020 and at the end of our study period (March 28, 2021). The national French attack rates are respectively estimated to be 5.7%, 8.8%, and 25.3% at these dates.

Basic reproductive number After adjusting for average weather over a period of one year in France, the estimate the national average basic reproductive number at 3.10 [2.95 ; 3.26].

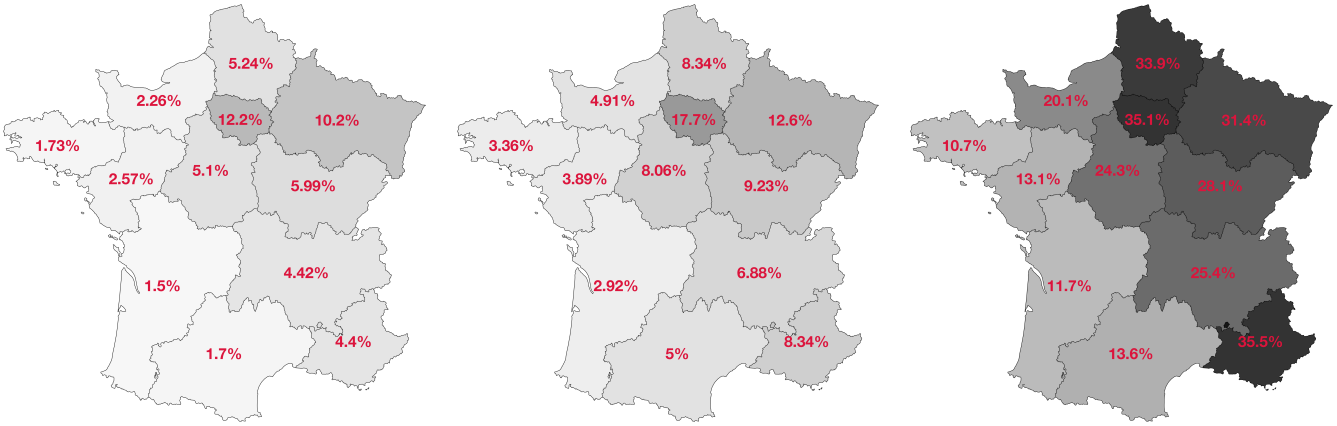

Figure 6: Model estimation for the proportion of naturally immunized individual in the population (deaths and vaccinated people not taken into account) on May 11$^{\text{th}}$, 2020 (left), on Oct. 5$^{\text{th}}$, 2020 (middle) and on March 28$^{\text{th}}$, 2021 (right).

	Modification of the transmission rate	
Covariate	**scale**	**increase/decrease**
NPI		
Lockdown 1 - delay of 7 days	-78% [-82% ; -74%]	↘
Post lockdown 1 - Phase 1	-54% [-56% ; -52%]	↘
Post lockdown 1 - Phase 2	-48% [-50% ; -47%]	↘
Lockdown 2 - delay of 7 days	-54% [-57% ; -49%]	↘
Lockdown 2 with opened shops	-51% [-53% ; -49%]	↘
Closing schools	- 7% [- 8% ; - 5%]	↘
Barrier gestures	-46% [-48% ; -44%]	↘
Curfew at 6PM	-30% [-33% ; -26%]	↘
Curfew at 8PM	-28% [-31% ; -25%]	↘
Closing bars & restaurants (ref. $W = 0$)	-10% [-13% ; - 8%]	↘
Closing bars & restaurants during summer (vs. ref.)	- 8% [-11% ; - 4%]	↘
during winter (vs. ref.)	-11% [-14% ; - 8%]	↘
Other factors		
100% of VOC circulating	22% [15% ; 28%]	↗
Weather effect during summer (ref. $W = 0$)	-22% [-24% ; -21%]	↘
during winter (ref. $W = 0$)	10% [9% ; 11%]	↗

Table 2: Estimation and 95% confidence intervals of the effects of seasonal weather conditions, proportion of VOC and NPI on the transmission rates. Model AIC = -1,388.

Effects of NPI on transmission rate Table 2 summarises the estimation of the effect of NPI on the transmission rates derived from the fixed effects of the model provided in Section 2.5. Regression residuals, fixed and random effects are given in Section 4.2 of Web Supplementary Materials. Figure 7 shows the corresponding individual fits. The very restrictive first lockdown (resp. the less restrictive second lockdown) decreases the transmission rate of $\sim 80\%$ (resp. $\sim 55\%$). The impact of school closures is around 7% and the bar and restaurants closures around 10%. The barrier gestures decrease the transmission of $\sim 46\%$. One can see that the values between the curfew at 6PM and the curfew at 8PM (around 30%) are very close and not statistically different (3% [-2% ; 8%]). We demonstrate that all NPI in this analysis decrease the transmission and have

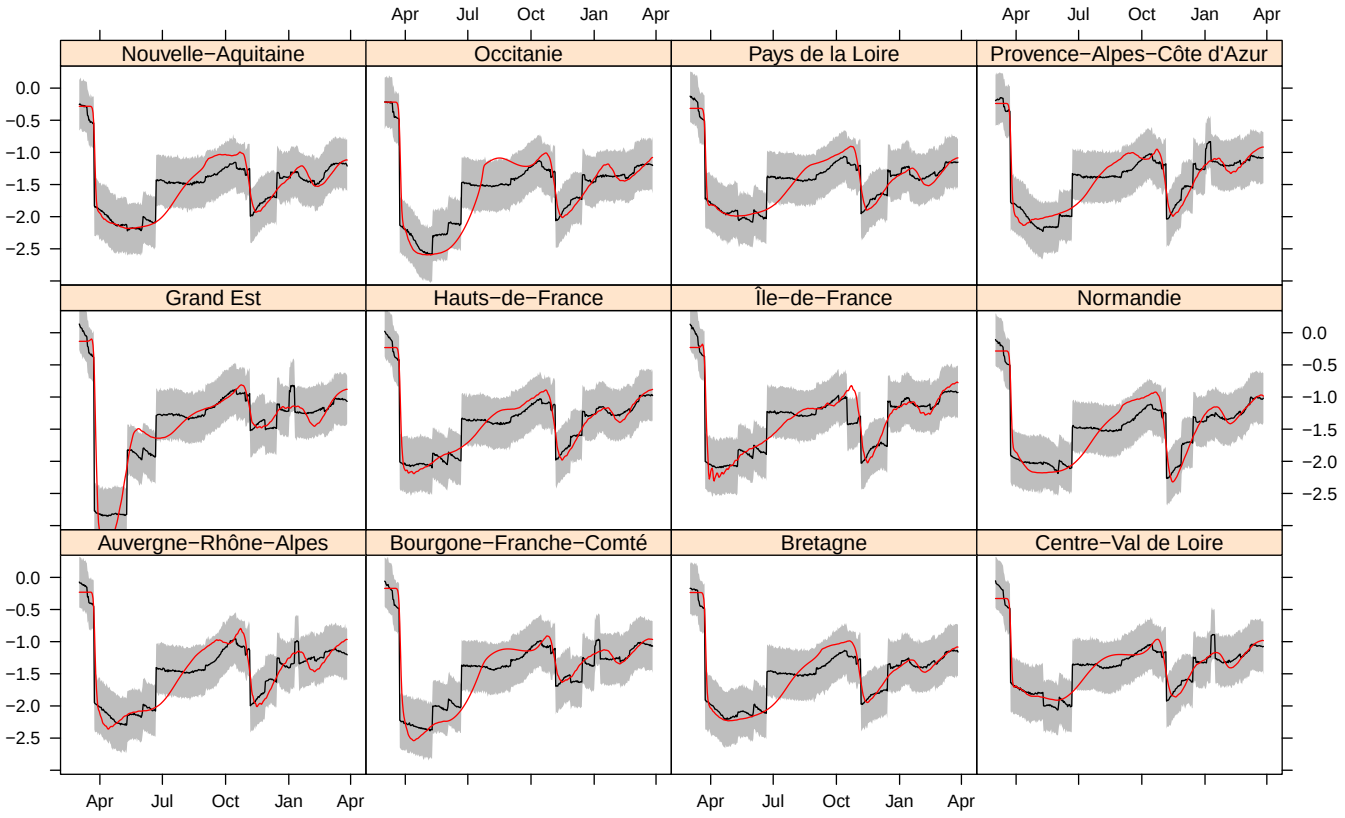

Figure 7: Results of regression model on the transmission rate obtained with the population based Kalman filter. Individual fits of regression model. Random effects on $\log(b_{init})$, b_{l1}, b_{l2} and $b_{curf6PM}$.

an effect significantly different from zero.

Effect of weather The seasonal weather conditions variable being unitless, the interpretation of its estimated effect should be made by comparing a value to a reference, which is set to be the average weather in France. We find that transmission is in average significantly increased by 10% during winter period and significantly decreased by 22% during summer period compared to average weather. Figure 8-top presents the estimated effect of the seasonal weather conditions over the period of the study by region. The estimated interaction between the closure of bars and restaurants and the seasonal weather conditions is statistically significant (p=0.037) but is also complex to interpret. We show that although closing bars and restaurants always has a significant effect of decreasing the transmission, it is slightly more effective in winter (11% [8%;14%] decrease) rather than in summer (8% [4%;11%] decrease). Figure 8-Middle presents the estimated effect of bars and restaurants closure for the 12 regions of interest over the study period and its interaction with weather conditions. The weather conditions impacts the estimation, with a stronger effect on transmission reduction in Northern regions compared to Southern ones, and also a stronger effect observed in the 2021 spring due to more favorable weather.

Effect of VOC Concerning the effect of the three VOC, our results show that an increase from 0 to 100% of the proportion would increase the transmission rate by an estimated 22% [15% ; 28%]. Figure 8-bottom presents how the increased transmissibility varied over time due to proportion of VOC.

4 Discussion

In this paper, we use an innovative method to infer the transmission rate over time from hospitalization data and estimate the effect of multiple NPI, weather and VOC on it. We showed that all NPI considered have a significant and independent effect on the transmission rate. We additionally demonstrated a strong effect of weather condition, decreasing the transmission in summer period and increasing it in winter period, and an increased transmissibility of VOC.

Concerning the transmission rate, interpretation of our results is conditional on the mechanistic model illustrated in Figure 4, and careful attention must be given to the parameters set from the scientific literature and detailed in Table 1. This model does not take into account the age-structure of the population on the contrary to the works of [3, 22] using French data. However, we believe that using a population approach accounts for different intrinsic characteristics such as

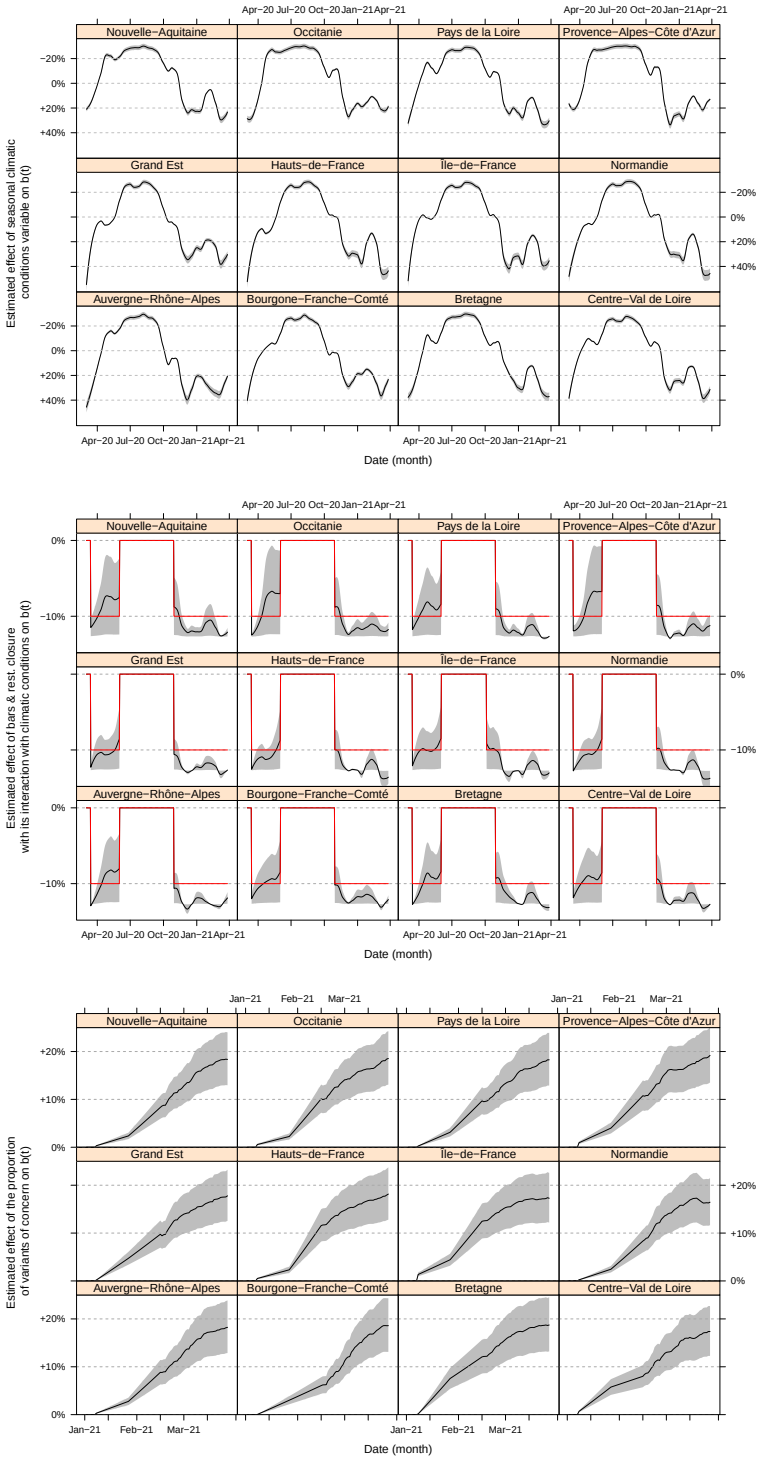

Figure 8: Top: Estimated effect of seasonal weather conditions and its 95% confidence band for the 12 regions of interest over the study period using the global average value over the period as reference for comparison. Middle: Estimated effect of bars and restaurants closure for the 12 regions of interest over the study period. In red, the main effect of -10%. In black and grey, the effect with the interaction with weather conditions and its 95% confidence band. Bottom: Estimated effect of VOC appearance and its 95% confidence band from January 1, 2021.

population density, age structure, transportation habits, etc. that influence disease spread through each region-specific transmission rates. Other drawback of the SEIRAH model is that it does not take into account the waning immunity or travels between regions. It could be included in the model by slightly modifying the model but it will imply to further fix more parameters without adding additional information from extra data, specially on inter-regional travels. For this reason we did not proceed to this development. Concerning the available data, the harm of using hospitalization data is its lack of exhaustiveness given that hospital services declaring cases may vary over time. However, one may think that this variation is less important than in other sources of data such as daily number of new confirmed cases for which the testing policy greatly changed since the beginning of the epidemics.

Interestingly, although the models were different and the data not fully identical, our results were comparable in terms of attack rates with existing modeling works [51] and sero-prevalence study [52, 53, 54]. A comparison is available in Web Supplementary material 4.1. Overall, our estimate tend to be slightly higher by less than 5% than other estimation and seroprevalence studies. This is presumably due to the strong assumption that there is no waning immunity in our model. Indeed, we assume that all individuals infected once have a high enough titer of antibody response to be tested systematically positive in seroprevalence studies. In terms of reproductive numbers, we found 3.10 [2.95 ; 3.26] (after removing the weather impact using the regression model) compared to 3.18 [3.09 ; 3.24] in Di Domenico et al. [22] and 2.90 [2.81 ; 3.01] in Salje et al. [3]. These comparison allow us to validate the estimation strategy which has the great advantage to be able to well estimate the transmission rate without any *a priori* on the shape of the transmission. Furthermore, the computational times are very interesting (few minutes for the full estimation on a classical work station without code optimization).

We restricted our analysis to the estimation of ten NPI effects. More variables could have been added, e.g. partial interventions (only applied for example in large cities in regions highly affected by COVID-19) – but those would be inconsistent with our model defined at a regional level. Other potential variables suffer from inconsistent definitions. Working from home is a good example. First, the adherence to this intervention (which can vary due to employee fatigue, organisation difficulties and the absence of legal constraint) is difficult to evaluate. Second, regarding quantitative indicators, the DARES carried out surveys for companies with 10 or more employees and showed that the mean percentage of remote work is around 29.6% [18.8%, 40.4%] with the max value during the two lockdowns [55]. However, this indicator account for working from home and paid vacation which may be very different in terms of behaviours. Hence by omitting the working from home in

the model, its effect is captured by the effect of lockdown and barrier gestures. It is very difficult to identify its proper effect because it fluctuates over time and according to the geographical areas without valid measurement of it. We also had to make a certain number of modeling assumptions. For some interventions, such as lockdowns, we considered a 7-days delay of implementation (which is visible on the transmission rates obtained with the Kalman filter). This choice is validated by other studies such as Dehning et al. [20], in which the delay can be up to 15 days. However, we investigated its impact on our regression adjustment. Not considering the delay of 7 days greatly deteriorates the fits, see Section 4.2 of Web Supplementary Materials for more details. Considering our choice to model the effect of the partial opening of schools during the May 11$^{\text{th}}$ to July 4$^{\text{th}}$ 2020 period as equal to 70% of the effect of a full closure could be considered as an oversimplification. Over this period following the first lockdown, schools reopened very gradually according to three different phases and school attendance growth was even more progressive. Up to June 2$^{\text{nd}}$, strong disparities in openings and attendance existed among regions, with a average of only 30% of pupils under 12 years old attending. Other levels of secondary schools ("collèges" and "lycées") reopened in early June and progressively received more students until summer vacations early July. Not to risk facing identifiability issues, we chose not to differentiate regions or phases of reopening and a ratio of 0.7 of closure effect was applied for all regions. Regarding the effect of curfews, we failed at demonstrating a statistically significant difference between 6PM and 8PM. This is maybe due to an identifiability issue as the 8PM curfew was in place during less than 3 weeks in many regions. An alternative explanation could be that both curfews were impacting globally social gathering the same way. As for example, both are preventing most of private dinners and parties (or at least reducing drastically the number of guests).

The way meteorological factors were taken in account into the model is debatable. We made the choice to use an index based on only 3 factors (temperature, absolute and relative humidity) gather into a single index based on a sum of normalized squared variations around the center of intervals defined by [34] for each of the three factors [35] and considered only its seasonal variation trough smoothing. The main advantage was to reduce the number of parameters to estimate in a linear model. Nonetheless, even if the study introducing the index presents elements suggesting a positive correlation with number of new cases, the influence of meteorological factors on all aspects of the epidemic remains a highly discussed topic. Recent reviews of literature published in late 2020 about links between COVID-19 and environmental factors [56, 57] reveal there is a growing body of literature on effects of meteorological factors on COVID-19. Results about the effect of temperature and humidity on transmission does not appear consistent across reported studies as

suggested in [31], but the reasons could lie in the large variety of geographical and climatic areas and seasonal time periods covered. They also report that wind speed and solar radiations effects are also studied, two variables we did not not take into consideration in our study. All in all, we saw a clear point to the inclusion of a variable to account for important weather variations around the year in France as a confounder for other variables. Section 4.2 of Web Supplementary Materials explored simpler models, all deteriorate the fits – without the variable modeling seasonal weather conditions and without the interaction between bars and restaurants closures and weather conditions. However, we remain extremely cautious about the interpretation of the estimated effects (main and interaction with bars and restaurants closure) and, beyond, of associated mechanisms. Finally, we consider an interaction between seasonal weather conditions and bar and restaurants closure justified by the use of terraces (which have been expanded in many places since the beginning of the pandemic). Of course other interactions can be considered and more complex models can be written but it may lead to over-fitting.

Looking back at the original data, the variant 20I/501Y.V1 (Alpha) appears to have always been predominant (over 90% at any time) compared to the two other VOC (Beta and Gamma) considered in French non-insular territory [39]. Widely variable estimations of transmissibility increase for the variant 20I/501Y.V1 (Alpha) have been proposed in the scientific literature, ranging from 29% to 90% [58, 59, 60, 61, 62] the lower values being in line with our own findings. The higher estimates could be partially explained because that new VOC appeared at the beginning of winter in England and in the USA. This may lead to a confusion between weather condition and VOC increased transmissibility in other studies, as mentionned by Campbell et al. [59]. We tested this assumption by removing the weather condition from our model and found an increased transmissibility of 43%, see Section 4.2 of Web Supplementary Material. We retained the weather condition in our model as it greatly improved the fit.

Altogether, this work is one of the first attempt to evaluate retrospectively the effect of multiple NPI over a period of one-year of COVID-19 epidemics. On top of applying novel methodology to an up-to-date application, this work could be extended to generate "what-if" scenarios and help in implementation of NPI for potential future epidemic waves.

Acknowledgements

The authors thank the `opencovid-19` initiative for their contribution in opening the data used in this article. This work is supported in part by Inria Mission COVID19, project GESTEPID. The

authors thank Linda Witkop, Jane Heffernan, Quentin Clairon, Thomas Ferté and Maria Pietro for constructive discussions about this work.

Author contributions

AC, BPH, PM, MP and RT designed the study. LL and CV analyzed the data. AC, PM and MP implemented the software code. AC, BPH and MP interpreted the results. AC, BPH, LL, PM and MP wrote the manuscript.

References

[1] Zunyou Wu and Jennifer M. McGoogan. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72,314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*, 323(13):1239–1242, 2020.

[2] Nathanael Lapidus, Juliette Paireau, Daniel Levy-Bruhl, Xavier de Lamballerie, Gianluca Severi, Mathilde Touvier, Marie Zins, Simon Cauchemez, Fabrice Carrat, and Sapris-Sero Study Group. Do not neglect SARS-CoV-2 hospitalization and fatality risks in the middle-aged adult population. *Infectious diseases now*, 51(4):380–382, 2021.

[3] Henrik Salje, Cécile Tran Kiem, Noémie Lefrancq, Noémie Courtejoie, Paolo Bosetti, Juliette Paireau, Alessio Andronico, Nathanaël Hozé, Jehanne Richet, Claire-Lise Dubost, Yann Le Strat, Justin Lessler, Daniel Levy-Bruhl, Arnaud Fontanet, Lulla Opatowski, Pierre-Yves Boëlle, and Simon Cauchemez. Estimating the burden of SARS-CoV-2 in France. *Science*, 369 (6500):208–211, 2020.

[4] Frederick J Angulo, Lyn Finelli, and David L Swerdlow. Estimation of US SARS-CoV-2 Infections, Symptomatic Infections, Hospitalizations, and Deaths Using Seroprevalence Surveys. *JAMA network open*, 4(1):e2033706, 2021.

[5] Annemarie B Docherty, Ewen M Harrison, Christopher A Green, Hayley E Hardwick, Riinu Pius, Lisa Norman, Karl A Holden, Jonathan M Read, Frank Dondelinger, Gail Carson, Laura Merson, James Lee, Daniel Plotkin, Louise Sigfrid, Sophie Halpin, Clare Jackson, Carrol Gamble, Peter W Horby, Jonathan S Nguyen-Van-Tam, Antonia Ho, Clark D Russell, Jake Dunning, Peter JM Openshaw, J Kenneth Baillie, and Malcolm G Semple. Features of 20

133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ*, 369:m1985, 2020.

[6] Terry C. Jones, Guido Biele, Barbara Mühlemann, Talitha Veith, Julia Schneider, Jörn Beheim-Schwarzbach, Tobias Bleicker, Julia Tesch, Marie Luisa Schmidt, Leif Erik Sander, Florian Kurth, Peter Menzel, Rolf Schwarzer, Marta Zuchowski, Jörg Hofmann, Andi Krumbholz, Angela Stein, Anke Edelmann, Victor Max Corman, and Christian Drosten. Estimating infectiousness throughout SARS-CoV-2 infection course. *Science*, 2021.

[7] Peng Wu, Fengfeng Liu, Zhaorui Chang, Yun Lin, Minrui Ren, Canjun Zheng, Yu Li, Zhibin Peng, Yin Qin, Jianxing Yu, Mengjie Geng, Xiaokun Yang, Hongting Zhao, Zhili Li, Sheng Zhou, Lu Ran, Benjamin J Cowling, Shengjie Lai, Qiulan Chen, Liping Wang, Tim K Tsang, and Zhongjie Li. Assessing asymptomatic, pre-symptomatic and symptomatic transmission risk of SARS-CoV-2. *Clinical Infectious Diseases*, 2021. doi: 10.1093/cid/ciab271.

[8] Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, and Helen Tatlow. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, 5(4):529–538, 2021.

[9] Nicholas G Davies, Adam J Kucharski, Rosalind M Eggo, Amy Gimma, W John Edmunds, and Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*, 5(7):e375–e385, 2020.

[10] Kate M Bubar, Kyle Reinholt, Stephen M Kissler, Marc Lipsitch, Sarah Cobey, Yonatan H Grad, and Daniel B Larremore. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Science*, 371(6532):916–921, 2021.

[11] Sen Pei, Sasikiran Kandula, and Jeffrey Shaman. Differential effects of intervention timing on COVID-19 spread in the United States. *Science advances*, 6(49):eabd6370, 2020.

[12] You Li, Harry Campbell, Durga Kulkarni, Alice Harpur, Madhurima Nundy, Xin Wang, Harish Nair, and the Usher Network for COVID-19 Evidence Reviews (UNCOVER) group. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. *The Lancet Infectious Diseases*, 21(2):193–202, 2021.

[13] Nicolas Banholzer, Eva van Weenen, Adrian Lison, Alberto Cenedese, Arne Seeliger, Bernhard Kratzwald, Daniel Tschernutter, Joan Puig Salles, Pierluigi Bottrighi, Sonja Lehtinen, Stefan Feuerriegel, and Werner Vach. Estimating the effects of non-pharmaceutical interventions on the number of new infections with COVID-19 during the first epidemic wave. *PLoS one*, 16 (6):e0252827, 2021.

[14] Nazrul Islam, Stephen J Sharp, Gerardo Chowell, Sharmin Shabnam, Ichiro Kawachi, Ben Lacey, Joseph M Massaro, Ralph B D'Agostino, and Martin White. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. *BMJ*, 370:m2743, 2020.

[15] Solomon Hsiang, Daniel Allen, Sébastien Annan-Phan, Kendon Bell, Ian Bolliger, Trinetta Chong, Hannah Druckenmiller, Luna Yue Huang, Andrew Hultgren, Emma Krasovich, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, 584(7820): 262–267, 2020.

[16] Yang Liu, Christian Morgenstern, James Kelly, Rachel Lowe, CMMID COVID-19 Working Group, and Mark Jit. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *BMC Medicine*, 19(1):1–12, 2021.

[17] Nils Haug, Lukas Geyrhofer, Alessandro Londei, Elma Dervic, Amélie Desvars-Larrive, Vittorio Loreto, Beate Pinior, Stefan Thurner, and Peter Klimek. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour*, 4(12):1303–1312, 2020.

[18] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Helen Coupland, Thomas A Mellan, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, Mélodie Monod, Imperial College COVID-19 Response Team, Azra C Ghani, Christl A Donnelly, Steven Riley, Michaela A C Vollmer, Neil M Ferguson, Lucy C Okell, and Samir Bhatt. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261, 2020.

[19] Moritz U. G. Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M. Pigott, Louis du Plessis, Nuno R. Faria, Ruoran Li, William P. Hanage, John S. Brownstein, Maylis Layan, Alessandro Vespignani, Huaiyu Tian, Christopher Dye, Oliver G. Pybus, and Samuel V. Scarpino. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490):493–497, 2020.

[20] Jonas Dehning, Johannes Zierenberg, F Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*, 369(6500), 2020.

[21] Jan M. Brauner, Sören Mindermann, Mrinank Sharma, David Johnston, John Salvatier, Tomáš Gavenčiak, Anna B. Stephenson, Gavin Leech, George Altman, Vladimir Mikulik, Alexander John Norman, Joshua Teperowski Monrad, Tamay Besiroglu, Hong Ge, Meghan A. Hartwick, Yee Whye Teh, Leonid Chindelevitch, Yarin Gal, and Jan Kulveit. Inferring the effectiveness of government interventions against COVID-19. *Science*, 371(6531), 2021.

[22] Laura Di Domenico, Giulia Pullano, Chiara E Sabbatini, Pierre-Yves Boëlle, and Vittoria Colizza. Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. *BMC Medicine*, 18(1):1–13, 2020.

[23] Dan Simon. *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches*. Wiley–Blackwell, 2006.

[24] Alain Bensoussan. *Estimation and Control of Dynamical Systems*. Interdisciplinary Applied Mathematics. Springer, 2018.

[25] Francisco Arroyo-Marioli, Francisco Bullano, Simas Kucinskas, and Carlos Rondón-Moreno. Tracking R of COVID-19: A new real-time estimation using the Kalman filter. *PLOS ONE*, 16(1):e0244474, 2021.

[26] Shariful Islam, Enamul Hoque, and Mohammad Ruhul Amin. Integration of Kalman filter in the epidemiological model: a robust approach to predict COVID-19 outbreak in Bangladesh. *medRxiv*, 2020. doi: 10.1101/2020.10.14.20212878.

[27] Bisong Hu, Wenqing Qiu, Chengdong Xu, and Jinfeng Wang. Integration of a Kalman filter in the geographically weighted regression for modeling the transmission of hand, foot and mouth disease. *BMC Public Health*, 20(1):479, 2020.

[28] Annabelle Collin, Mélanie Prague, and Philippe Moireau. Estimation for dynamical systems using a population-based Kalman filter – Applications to pharmacokinetics models. *Archive ouverte HAL*, 2020.

[29] Simon J Julier and Jeffrey K Uhlmann. Reduced Sigma Point Filters for the Propagation of Means and Covariances Through Nonlinear Transformations. In *In Proceedings of the 2002 American Control Conference*, volume 2, pages 887–892, 2002.

[30] French government's action summary website. https://www.gouvernement.fr/info-coronavirus/les-actions-du-gouvernement.

[31] Benjamin F Zaitchik, Neville Sweijd, Joy Shumake-Guillemot, Andy Morse, Chris Gordon, Aileen Marty, Juli Trtanj, Juerg Luterbacher, Joel Botai, Swadhin Behera, et al. A framework for research linking weather, climate and COVID-19. *Nature Communications*, 11(1):1–3, 2020.

[32] Xavier Rodó, Adrià San-José, Karin Kirchgatter, and Leonardo López. Changing climate and the COVID-19 pandemic: more than just heads or tails. *Nature Medicine*, 27:576–579, 2021.

[33] Xiaoyue Liu, Jianping Huang, Changyu Li, Yingjie Zhao, Danfeng Wang, Zhongwei Huang, and Kehu Yang. The role of seasonality in the spread of COVID-19 pandemic. *Environmental research*, 195:110874, 2021.

[34] Qasim Bukhari and Yusuf Jameel. Will coronavirus pandemic diminish by summer? *Available at SSRN 3556998*, 2020.

[35] Alix Roumagnac, Eurico de Carvalho Filho, Raphaël Bertrand, Anne-Kim Banchereau, and Guillaume Lahache. Étude de l'influence potentielle de l'humidité et de la température dans la propagation de la pandémie COVID-19. *Médecine de Catastrophe-Urgences Collectives*, 5 (1):87–102, 2021.

[36] World Health Organization. COVID-19 Weekly Epidemiological Update – 25 February 2021.

[37] Santé Publique France. COVID-19 – Point épidémiologique hebdomadaire du 28 Janvier 2021, .

[38] Santé Publique France. COVID-19 – Point épidémiologique hebdomadaire du 11 Mars 2021, .

[39] Données de laboratoires pour le dépistage : Indicateurs sur les variants (SI-DEP). https://www.data.gouv.fr/en/datasets/, .

[40] Données relatives aux personnes vaccinées contre la COVID-19. https://www.data.gouv.fr/fr/datasets/donnees-relatives-aux-personnes-vaccinees-contre-la-covid-19-1/, .

[41] Chaolong Wang, Li Liu, Xingjie Hao, Huan Guo, Qi Wang, Jiao Huang, Na He, Hongjie Yu, Xihong Lin, An Pan, Sheng Wei, and Tangchun Wu. Evolving Epidemiology and Impact of Non-pharmaceutical Interventions on the Outbreak of Coronavirus Disease 2019 in Wuhan, China. *medRxiv*, 2020. doi: 10.1101/2020.03.03.20030593.

[42] Mélanie Prague, Linda Wittkop, Annabelle Collin, Quentin Clairon, Dan Dutartre, Philippe Moireau, Rodolphe Thiebaut, and Boris Pierre Hejblum. Population modeling of early COVID-19 epidemic dynamics in French regions and estimation of the lockdown impact on infection rate. *medRxiv*, 2020. doi: 10.1101/2020.04.21.20073536.

[43] Jingjing He, Yifei Guo, Richeng Mao, and Jiming Zhang. Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis. *Journal of medical virology*, 93 (2):820–830, 2021.

[44] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, 368(6490):489–493, 2020.

[45] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9):577–582, 2020.

[46] Muge Cevik, Matthew Tate, Ollie Lloyd, Alberto Enrico Maraolo, Jenna Schafers, and Antonia Ho. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe*, 2020.

[47] J.-F. Delfraissy, L. Atlani Duault, D. Benamouzig, L. Bouadma, S. Cauchemez, F. Chauvin, A. Fontanet, A. Hoang, D. Malvy, and Y. Yazdanpanah. Une deuxième vague entrainant une situation sanitaire critique. In *Note du Conseil scientifique COVID-19*, 2020.

[48] Simon J Julier and Jeffrey K Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Signal Processing, Sensor Fusion, and Target Recognition VI*, volume 3068, pages 182–193, 1997.

[49] Marc Lavielle. *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC press, 2014.

[50] Jonathan Roux, Clément Massonnaud, and Pascal Crépey. COVID-19: One-month impact of the French lockdown on the epidemic burden. *medRxiv*, 2020. doi: 10.1101/2020.04.22.20075705.

[51] Nathanaël Hozé, Juliette Paireau, Nathanaël Lapidus, Cécile Tran Kiem, Henrik Salje, Gianluca Severi, Mathilde Touvier, Marie Zins, Xavier de Lamballerie, Daniel Lévy-Bruhl, Fabrice Carrat, and Simon Cauchemez. Monitoring the proportion of the population infected by SARS-CoV-2 using age-stratified hospitalisation and serological data: a modelling study. *The Lancet Public Health*, 6(6):e408–e415, 2021.

[52] J. Warszawski, N Bajos, L. Meyer, X. de Lamballerie, and et al. ER 1167: In May 2020, 4.5% of the population 2020 of metropolitan France had developed antibodies against SARS-CoV-2 – the first results of the EpiCov national survey. Technical report, DREES - INSERM, 2020.

[53] Fabrice Carrat, Xavier de Lamballerie, Delphine Rahib, Hélène Blanché, Nathanael Lapidus, Fanny Artaud, Sofiane Kab, Adeline Renuy, Fabien Szabo de Edelenyi, Laurence Meyer, Nathalie Lydié, Marie-Aline Charles, Pierre-Yves Ancel, Florence Jusot, Alexandra Rouquette, Stéphane Priet, Paola Mariela Saba Villarroel, Toscane Fourié, Clovis Lusivika-Nzinga, Jerome Nicol, Stephane Legot, Nathalie Druesne-Pecollo, Younes Esseddik, Cindy Lai, Jean-Marie Gagliolo, Jean-François Deleuze, Nathalie Bajos, Gianluca Severi, Mathilde Touvier, Marie Zins, and the SAPRIS and SAPRIS-SERO study groups. Seroprevalence of SARS-CoV-2 among adults in three regions of France following the lockdown and associated risk factors: a multicohort study. *medRxiv*, 2020. doi: 10.1101/2020.09.16.20195693.

[54] Santé Publique France. COVID-19 – Point épidémiologique hebdomadaire du 32 Décembre 2020, .

[55] Déchiffrer le monde du travail pour éclaire le débat public (DARES) - Principaux résultats des derniers mois - avril 2021. https://dares.travail-emploi.gouv.fr/publication/activite-et-conditions-demploi-de-la-main-doeuvre-pendant-la-crise-sanitaire-covid-19-1.

[56] Luis-Alberto Casado-Aranda, Juan Sánchez-Fernández, and María I Viedma-del Jesús. Analysis of the scientific production of the effect of COVID-19 on the environment: A bibliometric study. *Environmental Research*, 193:110416, 2021.

[57] Nayereh Rezaie Rahimi, Reza Fouladi-Fard, Rahim Aali, Ali Shahryari, Mostafa Rezaali, Yadollah Ghafouri, Mohammad Rezvani Ghalhari, Mahdi Asadi Ghalhari, Babak Farzinnia, Maria Fiore, et al. Bidirectional Association Between COVID-19 and the Environment: a Systematic Review. *Environmental Research*, page 110692, 2020.

[58] Mark S Graham, Carole H Sudre, Anna May, Michela Antonelli, Benjamin Murray, Thomas Varsavsky, Kerstin Kläser, Liane S Canas, Erika Molteni, Marc Modat, David A Drew, Long H Nguyen, Lorenzo Polidori, Somesh Selvachandran, Christina Hu, Joan Christina Hu Capdevila, COVID-19 Genomics UK (COG-UK) Consortium, Alexander Hammers, Andrew T Chan, Jonathan Wolf, Tim D Spector, Claire J Steves, and Sebastien Ourselin. Changes in symptomatology, reinfection, and transmissibility associated with the SARS-CoV-2 variant B.1.1.7: An ecological study. *The Lancet Public Health*, 6(5):e335–e345, 2021. doi: 10.1016/S2468-2667(21)00055-4.

[59] Finlay Campbell, Brett Archer, Henry Laurenson-Schafer, Yuka Jinnai, Franck Konings, Neale Batra, Boris Pavlin, Katelijn Vandemaele, Maria D Van Kerkhove, Thibaut Jombart, Oliver Morgan, and Olivier le Polain de Waroux. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Eurosurveillance*, 26(24):2100509, 2021.

[60] Nicole L Washington, Karthik Gangavarapu, Mark Zeller, Alexandre Bolze, Elizabeth T Cirulli, Kelly M Schiabor Barrett, Brendan B Larsen, Catelyn Anderson, Simon White, Tyler Cassens, Sharoni Jacobs, Geraint Levan, Jason Nguyen, Jimmy M Ramirez, Charlotte Rivera-Garcia, Efren Sandoval, Xueqing Wang, David Wong, Emily Spencer, Refugio Robles-Sikisaka, Ezra Kurzban, Laura D Hughes, Xianding Deng, Candace Wang, Venice Servellita, Holly Valentine, Peter De Hoff, Phoebe Seaver, Shashank Sathe, Kimberly Gietzen, Brad Sickler, Jay Antico, Kelly Hoon, Jingtao Liu, Aaron Harding, Omid Bakhtar, Tracy Basler, Brett Austin, Duncan MacCannell, Magnus Isaksson, Phillip G Febbo, David Becker, Marc Laurent, Eric McDonald, Gene W Yeo, Rob Knight, Louise C Laurent, Eileen de Feo, Michael Worobey, Charles Y Chiu, Marc A Suchard, James T Lu, William Lee, and Kristian G Andersen. Emergence and rapid transmission of SARS-CoV-2 B. 1.1. 7 in the United States. *Cell*, 184(10): 2587–2594.e7, 2021.

[61] Erik Volz, Swapnil Mishra, Meera Chand, Jeffrey C. Barrett, Robert Johnson, Lily Geidelberg, Wes R Hinsley, Daniel J Laydon, Gavin Dabrera, Áine O'Toole, Roberto Amato, Manon Ragonnet-Cronin, Ian Harrison, Ben Jackson, Cristina V. Ariani, Olivia Boyd, Nicholas J Loman, John T McCrone, Sónia Gonçalves, David Jorgensen, Richard Myers, Verity Hill, David K. Jackson, Katy Gaythorpe, Natalie Groves, John Sillitoe, Dominic P. Kwiatkowski, Seth Flaxman, Oliver Ratmann, Samir Bhatt, Susan Hopkins, Axel Gandy, Andrew Rambaut, and Neil M Ferguson. Transmission of SARS-CoV-2 Lineage B. 1.1. 7 in England: Insights

from linking epidemiological and genetic data. *medRxiv*, pages 1–12, 2021. doi: 10.1101/2020.12.30.20249034.

[62] Nicholas G Davies, Sam Abbott, Rosanna C Barnard, Christopher I Jarvis, Adam J Kucharski, James D Munday, Carl A B Pearson, Timothy W Russell, Damien C Tully, Alex D Washburne, Tom Wenseleers, Amy Gimma, William Waites, Kerry L M Wong, Kevin van Zandvoort, Justin D Silverman, CMMID COVID-19 Working Group1, COVID-19 Genomics UK (COG-UK) Consortium, Karla Diaz-Ordaz, Ruth Keogh, Rosalind M Eggo, Sebastian Funk, Mark Jit, Katherine E Atkins, and W John Edmunds. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, 372(6538):eabg3055, 2021.